

NO-0176 700

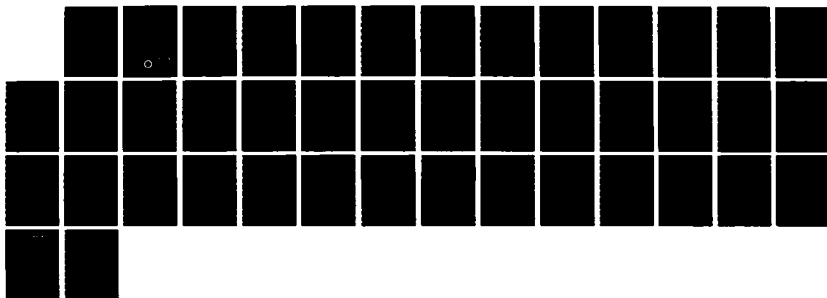
AN ANALYSIS OF BAYESIAN INFERENCE FOR NON-PARAMETRIC
REGRESSION(U) WASHINGTON UNIV SEATTLE DEPT OF
STATISTICS D D COX OCT 86 N00014-84-C-0169

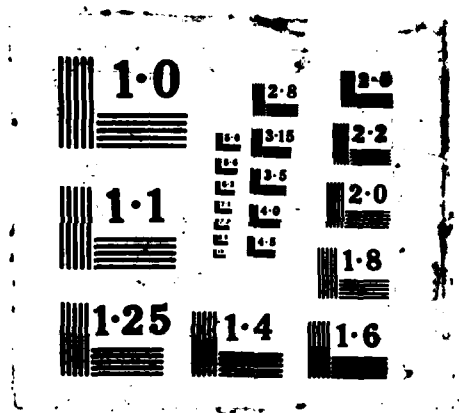
1/1

UNCLASSIFIED

F/8 12/1

NL





12

**AN ANALYSIS OF BAYESIAN INFERENCE
FOR NON-PARAMETRIC REGRESSION**

by

DENNIS D. COX

TECHNICAL REPORT No. 93

October 1986

Department of Statistics, GN-22

University of Washington

Seattle, Washington 98195 USA



**DTIC
ELECTE
FEB 11 1987
S E D**

AD-A176 788

DTIC FILE COPY

An Analysis of Bayesian Inference for Non-Parametric Regression

Dennis D. Cox*

Department of Statistics
University of Illinois
Urbana, IL 61801

on For	
DTIC TAB	<input checked="" type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	<input type="checkbox"/>
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

ABSTRACT

The observation model $y_i = \beta(i/n) + \varepsilon_i$, $1 \leq i \leq n$, is considered, where the ε 's are i.i.d. mean zero and variance σ^2 and β is an unknown smooth function. A Gaussian prior distribution is specified by assuming β is the solution of a high order stochastic differential equation. The estimation error $\delta = \beta - \hat{\beta}$ is analyzed, where $\hat{\beta}$ is the posterior expectation of β . Asymptotic posterior and sampling distributional approximations are given for $\|\delta\|^2$ when $\|\cdot\|$ is one of a family of norms natural to the problem. It is shown that the frequentist coverage probability of a variety of $(1 - \alpha)$ posterior probability regions tends to be larger than $1 - \alpha$, but will be infinitely often less than any $\varepsilon > 0$ as $n \rightarrow \infty$ with prior probability 1. A related continuous time signal estimation problem is also studied.



*This research was partially supported by the Office of Naval Research under contract number N00014-84-C-0169, and the National Science Foundation under grant number DMS-820-2560.

Key Words and Phrases: Bayesian inference; nonparametric regression; confidence regions; signal extraction; smoothing splines.

AMS-MOS Subject Classification (1980): Primary 62A15; Secondary 62G15, 62J99, 62E20, 62M99, 60G35.

This document has been approved for public release and its distribution is unlimited.

1. INTRODUCTION

In this article we consider Bayesian inference for a class of non-parametric regression models. Suppose we observe

$$Y_{ni} = \beta(t_{ni}) + \varepsilon_i \quad 1 \leq i \leq n \quad (1.1)$$

where $t_{ni} = i/n$, $\beta: [0,1] \rightarrow \mathbb{R}$ is an unknown smooth function, and $\varepsilon_1, \varepsilon_2, \dots$ are i.i.d. random errors with mean 0 and known variance $\sigma^2 < \infty$. The ε_{ni} are modeled as $N(0, \sigma^2)$. A Gaussian prior for β will now be specified. Let $m \geq 2$ and for some constants a_0, \dots, a_m with $a_m \neq 0$ let

$$L = \sum_{i=0}^m a_i D^i$$

be a constant coefficient linear differential operator ($D = d/dt$). Let ω denote a standard Gaussian white noise on $[0,1]$, which is formally the derivative of a standard Brownian motion $\langle W(t): 0 \leq t \leq 1 \rangle$. Let Σ be an $m \times m$ positive definite matrix and let B_1, \dots, B_m be boundary value operators of order $m-1$ or less, i.e., $B_i(\beta)$ a linear form in

$$\beta(0), \beta(1), D\beta(0), D\beta(1), \dots, D^{m-1}\beta(0), D^{m-1}\beta(1).$$

Assume that $Lg = 0$ and $B_i(g) = 0$, $1 \leq i \leq m$ imply $g = 0$. Then β is assumed to be the solution of the stochastic differential equation

$$\begin{aligned} L\beta &= \omega \\ \mathbf{B}(\beta) &\sim N(0, \Sigma) \end{aligned} \quad (1.2)$$

where $\mathbf{B}(\beta) = (B_1(\beta), \dots, B_m(\beta))$. Alternatively, one may express β through a simple stochastic integral

$$\beta(t) = \int_0^1 G(t,s) dW(s) + \beta_1(t)$$

where G is the Green's function for the operator L with homogeneous boundary conditions $B_i(\beta) = 0$, $1 \leq i \leq m$ (see sections 3.3 and 3.4 of Naimark, 1967) and $\beta_1(t)$ is a solution to the homogeneous differential equation ($L\beta_1 = 0$) with boundary conditions that agree with β ($B_i(\beta_1) = B_i(\beta)$, $1 \leq i \leq m$).

This specifies the prior for β . The Bayesian model thus defined is very similar to others that have appeared in the literature (Kimeldorf and Wahba, 1970a, 1970b; Wahba, 1978; Wecker and Ansley, 1983). More abstract models are discussed in section 2.

One of the attractive features of the Bayesian approach is that in principle one can solve virtually any statistical decision or inference problem. In particular, one can provide an accuracy assessment for $\hat{\beta}_n = E[\beta | Y_n]$ using posterior probability regions. For instance, letting $\|\cdot\|$ denote $L_2[0,1]$ norm, one can in principle determine a number Δ_n such that

$$P[\|\hat{\beta}_n - \beta\|^2 \leq \Delta_n | Y_n] = 0.95,$$

thus giving a 95% posterior probability bound on the L_2 -norm of the estimation error. A useful large sample approximation for Δ_n follows from results given below, namely

$$\Delta_n \sim \mu_n + 1.645 \tau_n$$

where

$$\mu_n \sim \pi^{-1} a_m^{-1/m} B(\frac{1}{2}m, 1 - \frac{1}{2}m)(\sigma^2/n)^{1-1/2m}$$

$$\tau_n^2 \sim \pi^{-1} a_m^{-1/m} B(\frac{1}{2}m, 2 - \frac{1}{2}m)(\sigma^2/n)^{2-1/2m}$$

and $B(x, y)$ is the beta function. (See (3.18), Theorem 3.1(a), (3.3), (3.1), and the remarks after Theorem 4.1.)

Non-Bayesians often find such Bayesian procedures attractive because as $n \rightarrow \infty$, the frequentist coverage probability of the Bayesian region tends to the posterior coverage probability in "typical" cases. It was my hope that this would hold in the nonparametric regression setting, thus providing a methodology for constructing large sample confidence regions. Unfortunately, the hoped for result is false in about the worst possible way, viz.

$$\liminf_{n \rightarrow \infty} P[\|\hat{\beta}_n - \beta\|^2 \leq \Delta_n | \beta] = 0, \text{ a.s.} \quad (1.3)$$

Thus, if one fixes a sample path from the Gaussian prior, then the frequentist coverage probability of the region $\{\beta: \|\beta - \hat{\beta}_n\|^2 \leq \Delta_n\}$ will infinitely often be arbitrarily small as $n \rightarrow \infty$, for almost all sample paths.

Nonetheless, for fixed n (large), the frequentist coverage probability is at least 95% for "most" sample paths in the following sense:

$$\lim_{n \rightarrow \infty} P[P[\|\hat{\beta}_n - \beta\|^2 \leq \Delta_n | \beta] \geq 0.95] = 1 - \alpha$$

where $1 - \alpha$ depends on m as indicated in the following table:

m	2	3	4	10	∞
$1 - \alpha$	89.4%	90.5%	91.2%	92.7%	95%

(see Corollary 3.4).

It will be indicated in section 5 that the "typical" settings wherein the frequentist coverage probability converges to the Bayesian posterior coverage probability depend critically on finite dimensionality of the parameter space. One would naturally conjecture that a result like (1.3) holds whenever the prior is infinite dimensional.

The procedure analyzed here differs from that advocated by Wahba (1983) and Wecker and Ansley (1983) in the following ways.

- (1) These authors use a "mixed effects" model wherein $\beta = \beta_0 + \beta_1$ and β_1 is given a proper prior of the type above but β_0 is unpriorized and lies in a finite dimensional space (or can be said to have an improper Lebesgue prior). The resulting regions are a mixture of confidence and posterior probability. I believe (1.3) will hold for this setting as well.
- (2) These authors only look at pointwise confidence intervals for $\beta(t)$ for some $t \in [0,1]$ rather than the global confidence regions treated here. I have not succeeded at analyzing the pointwise intervals, but I believe that (1.3) will hold for them as well. See section 5 for some relevant discussion.

- (3) We have assumed σ^2 is known, but I am sure that this is inconsequential.
- (4) These authors assume $\beta = b\beta_0$ where β_0 has a given prior and $b > 0$ is an unknown scale factor which is estimated. I do not know if our negative results still hold in this setting. In particular (1.3) depends on the law of the iterated logarithm fluctuations of the bias $E[\hat{\beta}_n - \beta \mid \beta]$ about its mean (Lemma 3.2). Such fluctuations undoubtedly impact the smoothing parameter estimation procedure of Wahba (1983), known as generalized cross validation. (See also Craven and Wahba (1979) and Speckman (1983).) Thus, an empirical Bayes approach may avoid this problem.
- (5) Wahba (1983) considers the true function β to be fixed and smoother than one generated by the Gaussian prior. I believe (1.3) may hold for many such smooth functions.

Section 2 presents an analysis of an abstract version of the Bayesian linear model. In section 3, a continuous time analog of (1.1) is investigated. In section 4, it is shown that the discrete model (1.1) can be well approximated by the continuous time model. Section 5 contains some concluding remarks.

2. THE ABSTRACT GAUSS-BAYES LINEAR MODEL

In this section, we formulate and solve a general version of the abstract linear model when the error vector is modeled as Gaussian with known covariance and a Gaussian prior is used for the parameter vector. Well known results about Gaussian measures on Banach spaces will be used, for which Kuo (1975) is an excellent reference; see also Kuelbs (1970, 1971), and Kallianpur (1971). *We will drop the subscript n throughout this section as we only consider a fixed prior and observation.*

Suppose ε is (modeled as) a mean 0 Gaussian random vector on a real separable Banach space \overline{Y} , and \underline{Y} is the generating Hilbert space (GHS) for the Gaussian measure $\underline{L}(\varepsilon)$ = the distribution of ε on \overline{Y} . (\underline{Y} is also known as the reproducing kernel Hilbert space for $\underline{L}(\varepsilon)$; see Kuelbs (1970), Kallianpur (1971), or Kuo (1975). Let β be a mean zero Gaussian random vector on a real separable Banach space $\overline{\Theta}$ with GHS Θ , and assume β and ε are independent. Now let

$$\overline{X}: \overline{\Theta} \rightarrow \underline{Y}$$

be a bounded linear operator, called the design operator, and suppose we observe

$$Y = \overline{X}\beta + \sigma\varepsilon,$$

where $\sigma > 0$ is known. We wish to estimate β from Y . To do this, we will show that a posterior distribution exists and characterize it.

As a remark, we note that \overline{Y} plays little role in what follows - it is Y that is important. If it were not for the technical problem that $\underline{L}(\varepsilon)$ will not live on Y when $\dim Y = \infty$, we could carry out the analysis using only Y with no mention of \overline{Y} . Kallianpur and Karandikar (1985) have an elegant approach to problems in this vein. Our approach here is in the more classical style using ordinary measure theory, for which it is necessary to keep \overline{Y} around.

Proposition 2.1. *There is a regular conditional posterior distribution $\underline{L}(\beta|Y)$, and in fact $\underline{L}(\beta|Y)$ is absolutely continuous w.r.t. the prior $\underline{L}(\beta)$, written $\underline{L}(\beta|Y) \ll \underline{L}(\beta)$.*

Proof. The assumption that the range $R(\overline{X}) \subseteq Y$ guarantees the sampling distribution $\underline{L}(Y|\beta) \ll \underline{L}(\sigma\varepsilon)$ by Theorem 3.1, p. 118 of Kuo (1972).

An elementary Fubini argument completes the proof. \square

As the inclusion map $\iota: \Theta \rightarrow \overline{\Theta}$ is continuous, it follows that the restriction X of \overline{X} to Θ is in $\underline{B}(\Theta, Y)$, the set of all bounded linear operators from $\Theta \rightarrow Y$. Let $X^* \in \underline{B}(Y, \Theta)$ denote the adjoint and put

$$U = X^*X \in \underline{B}(\Theta),$$

where $\underline{B}(\Theta) = \underline{B}(\Theta, \Theta)$. The description of the posterior is most easily accomplished using the singular value decomposition of X , given next.

Lemma 2.2. *There exist complete orthonormal systems (abbreviated c.o.n.s.) $\langle \psi_\nu: \nu=1, 2, \dots \rangle$ for Θ and $\langle \eta_\nu: \nu=1, 2, \dots \rangle$ for $\overline{R}(X)$, the closure of $R(X)$ in \underline{Y} , and constants $\langle \alpha_\nu^2: \nu=1, 2, \dots \rangle$ such that $U\psi_\nu = \alpha_\nu^2\psi_\nu$ for $\forall \nu$, and*

$$X\psi_\nu = \alpha_\nu\eta_\nu, \quad \forall \nu \quad (2.1)$$

$$X^*\eta_\nu = \alpha_\nu\psi_\nu, \quad \forall \nu \quad (2.2)$$

$$\Sigma \alpha_\nu^2 < \infty. \quad (2.3)$$

Proof. $X = \overline{X} \circ \iota$ and ι is a compact operator. Hence X , U and $V = XX^*$ are compact. U is also self adjoint and nonnegative definite on Θ , so there are nonnegative eigenvalues $\alpha_1^2, \alpha_2^2, \dots$ and eigenvectors ψ_1, ψ_2, \dots such that $\langle \psi_\nu \rangle$ is a c.o.n.s. for Θ , and $U\psi_\nu = \alpha_\nu^2\psi_\nu$ (Theorem 1.8, page 8 of Kuo, 1975). We assume w.l.o.g. that $\alpha_1^2 \geq \alpha_2^2 \geq \dots$. Put $\eta_\nu = \alpha_\nu^{-1}X\psi_\nu$ for $\alpha_\nu > 0$ and then (2.1) and (2.2) hold. Now we have $\beta = \sum_\nu \langle \beta, \psi_\nu \rangle_\Theta \psi_\nu$ with convergence $\underline{L}(\beta)$ a.s. in $\overline{\Theta}$, so $\overline{X}\beta = \sum_\nu \alpha_\nu \langle \beta, \psi_\nu \rangle_\Theta \eta_\nu$. As this latter series converges $\underline{L}(\beta)$ a.s. in \underline{Y} , and $\langle \beta, \psi_1 \rangle_\Theta, \langle \beta, \psi_2 \rangle_\Theta, \dots$ are i.i.d. $N(0,1)$, it follows that $\Sigma \alpha_\nu^2 < \infty$. Alternatively, one checks that V is the covariance of $X\beta$ in \underline{Y} , so trace $V = \Sigma \alpha_\nu^2 < \infty$. \square

Even though $\beta \in \Theta$ with probability 0, the "stochastic linear functional" $\beta \rightarrow \langle \theta, \beta \rangle_\Theta^\sim$ is defined a.s. as a measurable map $\overline{\Theta} \rightarrow \mathbb{R}$, for each

fixed $\theta \in \Theta$. The properties of these maps are given by Kuelbs (1971) and Kuo (1975), Lemma 4.7, p. 78. Similar remarks hold for ε and $\langle \varepsilon, \eta \rangle_{\underline{Y}}^{\sim}$, $\eta \in \underline{Y}$. Note that $\langle Y, \eta \rangle_{\underline{Y}}^{\sim} = \langle X\beta, \eta \rangle_{\underline{Y}}^{\sim} + \sigma \langle \varepsilon, \eta \rangle_{\underline{Y}}^{\sim}$, a.s. Since $X\beta \in \underline{Y}$, the term $\langle X\beta, \eta \rangle_{\underline{Y}}^{\sim}$ is defined as an ordinary linear function. We also have Karhunen-Loeve expansions: if e_1, e_2, \dots is any c.o.n.s. for Θ , then $\beta = \sum_{\nu} \langle \beta, e_{\nu} \rangle_{\Theta}^{\sim} e_{\nu}$ with convergence in $\overline{\Theta}$, almost surely. See Kuelbs (1971).

Lemma 2.3. Let $Y_{\nu} = \langle Y, \eta_{\nu} \rangle_{\underline{Y}}^{\sim}$ and $\beta_{\nu} = \langle \beta, \psi_{\nu} \rangle_{\Theta}^{\sim}$, $\nu = 1, 2, \dots$. Then

$$L(\beta_1, \beta_2, \dots | Y) = \otimes_{\nu} L(\beta_{\nu} | Y_{\nu}) = \otimes_{\nu} N((\sigma^2 + \alpha_{\nu}^2)^{-1} \alpha_{\nu} Y_{\nu}, [1 + (\alpha_{\nu}/\sigma)^2]^{-1})$$

i.e., the posterior distribution of β_1, β_2, \dots is that of independent normal random variables with the indicated means and variances.

Proof. Put $\varepsilon_{\nu} = \langle \varepsilon, \eta_{\nu} \rangle_{\underline{Y}}^{\sim}$. Then $\beta_1, \beta_2, \dots, \varepsilon_1, \varepsilon_2, \dots$ are i.i.d. $N(0, 1)$. Also $Y_{\nu} = \alpha_{\nu} \beta_{\nu} + \sigma \varepsilon_{\nu}$. The result follows from this and elementary calculations. \square

As the inclusion $\iota: \Theta \rightarrow \overline{\Theta}$ is continuous, every $\zeta \in \overline{\Theta}^*$, the dual of $\overline{\Theta}$, defines by restriction an element of Θ^* . If we identify $\Theta = \Theta^*$, then $\iota^* \in \underline{B}(\overline{\Theta}^*, \Theta)$ satisfies $\langle \zeta, \iota\theta \rangle_{\overline{\Theta}^*, \overline{\Theta}} = \langle \iota^* \zeta, \theta \rangle_{\Theta}$, $\forall \theta \in \Theta$. An inner product type notation is being used for the $(\overline{\Theta}^*, \overline{\Theta})$ duality pairing.

Theorem 2.4(a). The posterior $\underline{L}(\beta|Y)$ on $\overline{\Theta}$ is Gaussian with mean

$$\hat{\beta} = (\sigma^2 I + U)^{-1} X^* Y \quad \underline{L}(\beta, \epsilon) \text{ a.s.} \quad (2.4)$$

and covariance

$$V = \iota(I + \sigma^{-2}U)^{-1}\iota^*.$$

(b). $\delta = \beta - \hat{\beta}$ is independent of Y , and has a Gaussian distribution on $\overline{\Theta}$ with mean 0 and covariance V .

Remark. For (2.4) to make sense, we need to show X^*Y is defined a.s. and an element of Θ . Put

$$X^*Y = \sum_{\nu} Y_{\nu} X^* \eta_{\nu} = \sum_{\nu} \alpha_{\nu} Y_{\nu} \psi_{\nu}.$$

Note that $\Sigma E[(\alpha_{\nu} Y_{\nu})^2] = \Sigma \alpha_{\nu}^2 (\alpha_{\nu}^2 + \sigma^2) < \infty$, so the series on the right converges a.s. in Θ . Furthermore, we have for $\forall \theta \in \Theta$ that

$$\langle \theta, X^*Y \rangle_{\Theta} = \langle X\theta, Y \rangle_Y^{\sim} \quad \text{a.s.,}$$

as one can check from series expansions of both sides.

Proof. $\beta = \Sigma \beta_{\nu} \psi_{\nu}$ with a.s. convergence in $\overline{\Theta}$. Hence, if $\zeta \in \overline{\Theta}^*$,

$$\langle \zeta, \beta \rangle_{\overline{\Theta}^*, \overline{\Theta}} = \Sigma_{\nu} \beta_{\nu} \langle \zeta, \psi_{\nu} \rangle_{\overline{\Theta}^*, \overline{\Theta}} = \Sigma \beta_{\nu} \langle \iota^* \zeta, \psi_{\nu} \rangle_{\Theta}.$$

As $\Sigma_{\nu} \langle \iota^* \zeta, \psi_{\nu} \rangle_{\Theta}^2 = \|\iota^* \zeta\|_{\Theta}^2 < \infty$, it follows from Lemma 2.3 that

$\underline{L}(\langle \zeta, \beta \rangle_{\overline{\Theta}^*, \overline{\Theta}} | Y)$ is Gaussian with mean

$$\mu = \sum_{\nu} (\sigma^2 + \alpha_{\nu}^2)^{-1} \alpha_{\nu} Y_{\nu} \langle \iota^* \zeta, \psi_{\nu} \rangle_{\Theta}$$

and variance

$$\tau^2 = \sum_{\nu} [1 + (\alpha_{\nu}/\sigma)^2]^{-1} \langle \iota^* \zeta, \psi_{\nu} \rangle_{\Theta}^2$$

Note that

$$\begin{aligned} \langle \zeta, \iota(\sigma^2 I + U)^{-1} X^* Y \rangle_{\bar{\Theta}^*, \Theta} &= \langle (\sigma^2 I + U)^{-1} \iota^* \zeta, X^* Y \rangle_{\Theta} \\ &= \sum_{\nu} \langle (\sigma^2 I + U)^{-1} \iota^* \zeta, \psi_{\nu} \rangle_{\Theta} (\alpha_{\nu} Y_{\nu}) \\ &= \sum_{\nu} \alpha_{\nu} (\sigma^2 + \alpha_{\nu}^2)^{-1} \langle \iota^* \zeta, \psi_{\nu} \rangle_{\Theta} Y_{\nu} = \mu. \end{aligned}$$

Also $\langle \zeta, \iota(I + \sigma^2 U)^{-1} \iota^* \zeta \rangle_{\bar{\Theta}^*, \Theta} = \tau^2$. This completes the proof of (a).

Put $\delta_{\nu} = \langle \delta, \psi_{\nu} \rangle_{\Theta}^{\sim} = \beta_{\nu} - (\sigma^2 + \alpha_{\nu}^2)^{-1} \alpha_{\nu} Y_{\nu}$. It is easy to check that δ_{ν} is

Gaussian and independent of Y_{ν} , hence Y . The rest of (b) is easy. \square

Now we introduce a parametrized family of Hilbert spaces which is natural for the problem. For convenience, assume

$$\underline{R}(X^*) \text{ is dense in } \Theta \quad (2.5)$$

where $\underline{R}(X^*) = \underline{R}(U)$ is the range of X^* . Also, put

$$\gamma_{\nu} = \alpha_{\nu}^{-2}, \quad \varphi_{\nu} = \alpha_{\nu}^{-1} \psi_{\nu}.$$

For $\rho \in \mathbb{R}$ let Θ_{ρ} be the Hilbert space obtained by completing the set of finite-norm elements of Θ under the norm given by $\|\theta\|_{\rho} = \langle \theta, \theta \rangle_{\rho}^{\frac{1}{2}}$ where

$$\langle \theta_1, \theta_2 \rangle_\rho = \sum_\nu \gamma_\nu^\rho \langle \theta_1, U \varphi_\nu \rangle_\theta \langle \theta_2, U \varphi_\nu \rangle_\theta.$$

We collect some elementary facts about these spaces.

Proposition 2.5. *Suppose (2.5) holds, then*

- (a) $\langle \theta, \xi \rangle_0 = \langle \theta, U \xi \rangle_\theta$, and $\langle \varphi_\nu: \nu = 1, 2, \dots \rangle$ is a c.o.n.s. for θ_0 .
- (b) $\theta_1 = \theta$.
- (c) X extends to a Hilbert space isomorphism from θ_0 to the \underline{Y} closure of $\underline{R}(X)$.
- (d) If $\rho > \tau$ then $\theta_\rho \subseteq \theta_\tau$ with continuous inclusion.
- (e) Letting ι_ρ denote the inclusion $\theta \rightarrow \theta_\rho$ where $\rho \leq 1$, and identifying θ_ρ^* with θ_ρ in the usual way, we have $\iota_\rho^* = U^{1-\rho}$.
- (f) θ_ρ supports $\underline{L}(\beta)$ if and only if $\sum_\nu \gamma_\nu^{\rho-1} < \infty$.

Proof. (a) and (b) are elementary. For (c) note that $X\varphi_\nu = \eta_\nu$ and $\langle \theta, U \varphi_\nu \rangle_\theta = \langle X\theta, \eta_\nu \rangle_{\underline{Y}}$. For (d) note that $\|\theta\|_\rho^2 \leq (\max_\nu \gamma_\nu^{\rho-\tau}) \|\theta\|_\tau^2$, and the $\max_\nu \gamma_\nu^{\rho-\tau}$ is achieved at a finite, positive value by (2.3). For (e), note that $U^{1-\rho}\varphi_\nu = \gamma_\nu^{-(1-\rho)}\varphi_\nu$ and by comparison of the series defining the two sides of $\langle \iota_\rho \theta, \eta \rangle_\rho = \langle \theta, \iota_\rho^* \eta \rangle_1$, one obtains that $\iota_\rho^* \eta = \sum_\nu \gamma_\nu^{-(1-\rho)} \langle \eta, U \varphi_\nu \rangle_1 \varphi_\nu$ which gives the result.

Turning to (f), suppose $\sum \gamma_\nu^{\rho-1} < \infty$. Then the operator $D : \Theta \rightarrow \Theta$ given by $D\theta = \sum \gamma_\nu^{\rho-1/2} \langle \theta, \varphi_\nu \rangle_0 \varphi_\nu$ is Hilbert-Schmidt, so the norm $||| \cdot |||$ on Θ given by $||| \theta ||| = \|D\theta\|_1$ is a measurable norm (Exercise 17, page 59 of Kuo, 1975) and hence the completion of Θ under $||| \cdot |||$ supports $\underline{L}(\beta)$ (Theorem 4.1, page 63 of Kuo, 1975). One easily checks that $||| \cdot ||| = \|\cdot\|_\rho$. Conversely, suppose Θ_ρ supports $\underline{L}(\beta)$, and let us calculate the covariance operator V for $\underline{L}(\beta)$ on Θ_ρ . For $\theta \in \Theta$,

$$\begin{aligned} \langle \theta, V\theta \rangle_\rho &= E \langle \theta, \beta \rangle_\rho^2 = E \left[\left(\sum \gamma_\nu^\rho \langle \theta, \varphi_\nu \rangle_0 \langle \beta, \varphi_\nu \rangle_0 \right)^2 \right] \\ &= E \left[\left(\sum \gamma_\nu^{\rho-1/2} \langle \theta, \varphi_\nu \rangle_0 \beta_\nu \right)^2 \right] = \sum \gamma_\nu^{2\rho-1} \langle \theta, U\varphi_\nu \rangle_1^2 \\ &= \sum \gamma_\nu^\rho \langle \theta, \varphi_\nu \rangle_0 \langle U^{\rho-1}\theta, \varphi_\nu \rangle_0 = \langle \theta, U^{\rho-1}\theta \rangle_\rho, \end{aligned}$$

and since Θ is dense in Θ_ρ , we can identify $V = U^{\rho-1} = \iota_\rho \iota_\rho^*$. Now V must be a trace class operator (Theorem 2.3(a), page 29, and Definition 2.2, page 16 of Kuo, 1975), and the eigenvalues of V are $\{\gamma_\nu^{\rho-1}\}$, so $\sum \gamma_\nu^{\rho-1} < \infty$. \square

3. CONTINUOUS TIME ESTIMATION

In this section we consider a continuous time analog of the problem from section 1. It will be seen in the next section that the sequence of experiments with discrete observations can be sufficiently well approximated by a sequence of experiments with continuous observations that the limit theory of the latter is inherited by the former.

Suppose $\langle \beta(t) : 0 \leq t \leq 1 \rangle$ is the stochastic process of (1.2) and let $\langle \varepsilon(t) : 0 \leq t \leq 1 \rangle$ be a standard Gaussian white noise. Consider the sequence of experiments with observations

$$Y_n(t) = \beta(t) + n^{-1/2} \sigma \varepsilon(t) \quad , \quad 0 \leq t \leq 1 ,$$

where $\sigma > 0$ is a constant. We write

$$\lambda_n = \sigma_n^2 = \sigma^2 / n . \quad (3.1)$$

Now the GHS for $\underline{L}(\varepsilon)$ is

$$\underline{Y} = L_2[0,1] .$$

Θ , the GHS for $\underline{L}(\beta)$, is equal as a set to the Sobolev space

$$\begin{aligned} W_2^m = W_2^m[0,1] = \{ f : f \text{ maps } [0,1] \rightarrow \mathbb{R} , \text{ and} \\ f, Df, \dots, D^{m-1}f \text{ are absolutely continuous} \\ \text{with } D^m f \in L_2[0,1] \} . \end{aligned}$$

The Θ inner product is

$$\langle \theta, \zeta \rangle_{\Theta} = \mathbf{B}(\theta)' \Sigma^{-1} \mathbf{B}(\zeta) + \int_0^1 (L\theta)(L\zeta) .$$

One can show that $\|\cdot\|_{\Theta}$ is equivalent to $\|\cdot\|_{H_T^m}$ where

$$\|\theta\|_{H^m}^2 = \|D^m \theta\|_{L_2}^2 + \|\theta\|_{L_2}^2.$$

Now let $\bar{\Theta}$ be a Banach space which will support $L(\beta)$ and is continuously imbedded in L_2 (i.e., $\bar{\Theta} \subseteq L_2$ and has a stronger norm), e.g., $\bar{\Theta} = C^{m-1}$ or W_2^{m-1} . Let $\bar{X}: \bar{\Theta} \rightarrow L_2$ be the imbedding operator ($\bar{X}\theta = \theta$), so $Y = \bar{X}\beta + \sigma_n \varepsilon$. Then X , the restriction of \bar{X} to Θ , is the imbedding of Θ in L_2 . To identify X^* , start with

$$\int_0^1 (L\theta)(LX^*\eta) + B(\theta)' \Sigma^{-1} B(X^*\eta) = \int_0^1 \theta \eta$$

valid for $\theta \in \Theta$ and $\eta \in Y$. Apply integration by parts to the integral (as in section 1.5 of Naimark, 1967) and express the l.h.s. as

$$\int_0^1 (L^*L)(X^*\eta)\theta + \sum_{j=0}^{m-1} [C_{j0}(X^*\eta)D^j\theta(0) + C_{j1}(X^*\eta)D^j\theta(1)].$$

Here L^* is the formal adjoint of L . Also, $C_{ji}(\zeta)$ is a boundary value operator of order $2m-j$ whose highest order term is of the form $\pm a_m^2 D^{2m-j} \zeta(i)$. Thus $\zeta = X^*\eta$ can be obtained as a solution of the differential equation

$$L^*L\zeta = \eta$$

$$C_{ji}(\zeta) = 0, \quad 0 \leq j < m, \quad i = 1, 2.$$

To derive $\hat{\beta}_n = E[\beta | Y_n] = (\lambda_n I + U)^{-1} X^* Y_n$, it is convenient to use the following characterization: $\hat{\beta}_n$ is the unique $\beta \in \Theta$ such that for all $\theta \in \Theta$,

$$\lambda_n \langle \beta, \theta \rangle_{\Theta} + \langle X\beta, X\theta \rangle_Y = \langle Y_n, X\theta \rangle_Y.$$

If one writes this out and uses integration by parts, there results that $\hat{\beta}_n$ is the solution of the stochastic differential equation

$$(\lambda_n L^* L + 1) \hat{\beta} = Y_n$$

$$C_{ji}(\hat{\beta}) = 0, \quad 0 \leq j < m, \quad i = 1, 2.$$

The φ_ν 's and γ_ν 's are the eigenvectors and eigenvalues of the differential operator $L^* L$ with boundary conditions $C_{ji} = 0, 0 \leq j < m, i = 1, 2$. Since the highest order term of $C_{ji}(\zeta)$ is $\pm a_m^2 D^{2m-j} \zeta(i)$, the boundary conditions are regular by the argument on pp. 60-61 of Naimark (1967). By Theorem 2, pp. 64-65 of Naimark,

$$\gamma_\nu = a_m^2 (\pi \nu)^{2m} (1 + O(\nu^{-3/2})) \quad , \quad \nu \rightarrow \infty. \quad (3.2)$$

(Note that the eigenvalues have to be real and positive.) Put

$$C_k(\lambda, \rho) = \sum_\nu \gamma_\nu^\rho (1 + \lambda \gamma_\nu)^{-k}.$$

These quantities will have many uses. In particular note that

$$E \|\hat{\beta} - \beta\|_\rho^2 = \lambda_n C_1(\lambda_n, \rho)$$

$$E \|\hat{\beta} - E[\hat{\beta} | \beta]\|_\rho^2 = \lambda_n C_2(\lambda_n, \rho).$$

Using (3.2), one can show that if $-\frac{1}{2}m < \rho < 1 - \frac{1}{2}m$, then as $\lambda \rightarrow 0$

$$\begin{aligned} C_k(\lambda, \rho) &\sim \kappa \lambda^{-(\rho+1/2m)} \int_0^\infty x^{2m\rho} [1 + x^{2m}]^{-k} dx \\ &= \kappa \lambda^{-(\rho+1/2m)} B(\rho + 1/2m, k - \rho - 1/2m) \end{aligned} \quad (3.3)$$

where

$$\kappa = \pi^{-1} a_m^{-1/m}$$

and $B(u, v) = \Gamma(u)\Gamma(v)/\Gamma(u+v)$ is the beta function. The derivation for the asymptotic formula for $C_k(\lambda, \rho)$ follows as in Corollary 5.4 of Speckman (1984). The evaluation of the integral is given in Gradshteyn and Ryzhik (1965), formula 3.194.2, page 285.

Now $\Theta_0 = L_2$ and $\Theta_1 \cong W_2^m$, where \cong means equal as sets and with equivalent norms. From this it follows that

$$\Theta_\rho \cong W_2^{m\rho}, \quad 0 \leq \rho \leq 1. \quad (3.4)$$

The proof is most easily accomplished with the K -method of interpolation; see Theorem 3.4 of Cox (1986).

One of the reasons the norms $\|\cdot\|_\rho$ are so useful is because we can derive explicit limiting distribution results as $\lambda_n \rightarrow 0$ for the norm of the error vector $\delta_n = \beta - \hat{\beta}_n$.

Theorem 3.1. *Let $-1/2m < \rho < 1 - 1/2m$.*

(a) *Put*

$$\mu_n(\rho) = \lambda_n C_1(\lambda_n, \rho)$$

$$\tau_n^2(\rho) = 2\lambda_n^2 C_2(\lambda_n, 2\rho).$$

Then

$$\underline{L}((\|\delta_n\|_\rho^2 - \mu_n)/\tau_n) \Rightarrow N(0, 1)$$

as $n \rightarrow \infty$.

(b) Put

$$\Omega_n(\rho, \beta) = \lambda_n^2 \sum_{\nu} \gamma_{\nu}^{\rho+1} (1 + \lambda_n \gamma_{\nu})^{-2} (\beta_{\nu}^2 - 1)$$

$$\eta_n^2(\rho) = 2\lambda_n^2 [C_3(\lambda_n, 2\rho) + \lambda_n C_4(\lambda_n, 1+2\rho)].$$

Then

$$\underline{L}([\|\delta_n\|_{\rho}^2 - \mu_n - \Omega_n] / \eta_n | \beta) \Rightarrow N(0, 1)$$

as $n \rightarrow \infty$, $\underline{L}(\beta)$ a.s.

Remarks. In view of Theorem 2.4(b), part (a) gives the asymptotic posterior distribution for $\|\delta_n\|_{\rho}^2$. Part (b) gives the asymptotic sampling distribution of $\|\delta_n\|_{\rho}^2$. Note that Ω_n is random as it depends on β .

Proof. For (a), note that

$$\|\delta_n\|_{\rho}^2 = \sum_{\nu} \gamma_{\nu}^{\rho} \langle \delta_n, \varphi_{\nu} \rangle_0^2 = \sum_{\nu} \gamma_{\nu}^{\rho-1} \langle \delta_n, \psi_{\nu} \rangle_{\theta}^2$$

and $\langle \delta_n, \psi_1 \rangle_{\theta}, \langle \delta_n, \psi_2 \rangle_{\theta}, \dots$ are independent normal with mean 0 and variance $\text{Var} \langle \delta_n, \psi_{\nu} \rangle_{\theta} = \lambda_n \gamma_{\nu} (1 + \lambda_n \gamma_{\nu})^{-1}$. The result follows from Lindeberg's central limit theorem after suitable truncation of the infinite series for $\|\delta_n\|_{\rho}^2$.

Part (b) is somewhat more difficult. The conditional distribution of $\langle \delta_n, \psi_1 \rangle_{\theta}, \langle \delta_n, \psi_2 \rangle_{\theta}, \dots$ given β is that of independent normals with mean $\lambda_n \gamma_{\nu} \beta_{\nu} (1 + \lambda_n \gamma_{\nu})^{-1}$ and variance $\lambda_n \gamma_{\nu} (1 + \lambda_n \gamma_{\nu})^{-2}$. From this it follows that

$$E[\|\delta_n\|_\rho^2 | \beta] = \mu_n + \Omega_n \text{ and}$$

$$\text{Var}[\|\delta_n\|_\rho^2 | \beta] = \eta_n + 4\lambda_n^3 \sum_{\nu} \gamma_{\nu}^{1+2\rho} (\beta_{\nu}^2 - 1)(1 + \lambda_n \gamma_{\nu})^{-4}.$$

From (3.3) we have $\eta_n \approx \lambda_n^{2-2\rho-1/2m}$ while from Lemma 3.2 below, the second term above is $\underline{o}(\lambda_n^{2-2\rho-1/2m})$, a.s. Now the proof follows as in part

(a) \square

Lemma 3.2. *Let $\tau > 1/2m$ and k be such that*

$$k > \tau + 1/4m. \quad (3.5)$$

Put

$$\omega(n) = \sum_{\nu=1}^{\infty} (\lambda_n \gamma_{\nu})^{\tau} (1 + \lambda_n \gamma_{\nu})^{-k} (\beta_{\nu}^2 - 1).$$

Then for some $C \in (0, \infty)$,

$$\overline{\lim}_{n \rightarrow \infty} \lambda_n^{1/4m} (\log \log n)^{-1/2} \omega(n) = C, \quad \text{a.s.}$$

$$\underline{\lim}_{n \rightarrow \infty} \lambda_n^{1/4m} (\log \log n)^{-1/2} \omega(n) = -C, \quad \text{a.s.}$$

Proof. The first step is to show that there is a probability space carrying a probabilistic replica of $\omega(n)$, also denoted $\omega(n)$, and a standard Wiener process $\{W(t): t \geq 0\}$ such that the process given by

$$V(s) = - \int_0^s W(st) g(t) dt, \quad s \geq 0$$

$$g(t) = (d/dt)h(t^{2m}), \quad h(x) = x^{\tau}/(1+x)^k$$

satisfies

$$|\omega(n) - 2^{\frac{1}{2}} V((c \lambda_n)^{-1/2m})| = o(\lambda_n^{-1/4m} (\log \log n)^{\frac{1}{2}}) \quad (3.6)$$

almost surely, where

$$c = \alpha_m^2 \pi^{2m}.$$

Note that (3.5) guarantees convergence of the integral defining V .

Let $S_\nu = \sum_{j=1}^\nu 2^{-\frac{1}{2}} (\beta_j^2 - 1)$ and

$$D_1(n) = 2^{-\frac{1}{2}} \omega(n) = - \sum_\nu S_\nu [h(\lambda_n \gamma_{\nu+1}) - h(\lambda_n \gamma_\nu)]$$

where the last formula follows by partial summation. By Theorem 2.6.1, page 107 of Csörgö and Révész (1981), there is a probability space carrying a version of $\{S_\nu: \nu = 1, 2, \dots\}$ and a Wiener process $\{W(t): t \geq 0\}$ such that for all $\delta > 0$, as $\nu \rightarrow \infty$

$$|S_\nu - W(\nu)| = o(\nu^\delta) \quad \text{a.s.} \quad (3.7)$$

Let

$$D_2(n) = - \sum_\nu W(\nu) [h(\lambda_n \gamma_{\nu+1}) - h(\lambda_n \gamma_\nu)].$$

Note that

$$|h(\lambda \gamma_{\nu+1}) - h(\lambda \gamma_\nu)| = \lambda [\gamma_{\nu+1} - \gamma_\nu] |h'(\lambda \tilde{\gamma}_\nu)|$$

where $\gamma_\nu \leq \tilde{\gamma}_\nu \leq \gamma_{\nu+1}$. Utilizing (3.2) one can show that

$$\gamma_{\nu+1} - \gamma_\nu = c 2m \nu^{2m-1} (1 + o(\nu^{-\frac{1}{2}})). \quad (3.8)$$

Also,

$$|h'(\lambda \gamma_\nu)| \leq C_1 [(\lambda \nu^{2m})^{\tau-1} (1 + c \lambda \nu^{2m})^{-k} + (\lambda \nu^{2m})^\tau (1 + c \lambda \nu^{2m})^{-(k+1)}]$$

for some $C_1 \in (0, \infty)$. Collecting things together we have

$$|D_1(n) - D_2(n)| \leq C_2 \lambda_n^{-\delta/2m} \sum_{\nu} (\lambda_n \nu^{2m})^{1+(\delta-1)/2m} [(\lambda_n \nu^{2m})^{\tau-1} (1+c \lambda_n \nu^{2m})^{-k} + (\lambda_n \nu^{2m})^{\tau} (1+c \lambda_n \nu^{2m})^{-(k+1)}] \lambda_n^{1/2m} \quad (3.9)$$

where $C_2 \in (0, \infty)$ is random (from (3.7)) and depends on δ . Note that the summation above tends to a finite integral as $n \rightarrow \infty$. Hence, taking $\delta < 1/2$ gives that

$$|D_1(n) - D_2(n)| = o(\lambda_n^{-1/4m} (\log \log n)^{1/2}).$$

Next we show that $D_2(n)$ is suitably close to

$$D_3(n) = - \sum_{\nu} W(\nu) (c \lambda_n)^{1/2m} g((c \lambda_n)^{1/2m} \nu).$$

Two applications of the mean value theorem along with (3.8) yield

$$|D_2(n) - D_3(n)| \leq C \sum_{\nu} |W(\nu)| (\lambda_n \nu^{2m-1})^2 (1 + o(\nu^{-1/2})) |h''(c \lambda_n \nu^{2m} [1 + o(\nu^{-1})])|$$

for some $C \in (0, \infty)$, almost surely. If one uses the fact that $|W(\nu)| = o(\nu)$ a.s. and applies the same argument that was used on (3.9), it can be shown that the last expression is $o(1)$ a.s., which is $o(\lambda_n^{-1/4m} (\log \log n)^{1/2})$. (The assumption $\tau > 1/2m$ is needed here to guarantee that a summation converges to a finite integral.)

Now let $D_4(n) = V((c \lambda_n)^{-1/2m})$. Then

$$|D_3(n) - D_4(n)| \leq (c \lambda_n)^{-1/2m} \sum_{\nu} Z_{\nu} d_{n\nu} \quad (3.10)$$

where

$$Z_\nu = \sup_{\nu \leq t \leq \nu+1} |W(t) - W(\nu)|$$

is a sequence of i.i.d random variables and

$$d_{n\nu} = |g'([c\lambda_n]^{1/2m} [\nu + \alpha_{n\nu}])|$$

for some $\alpha_{n\nu} \in [0,1]$. Now $P[Z_\nu \geq y] = 4P[W(1) \geq y]$, so as $\nu \rightarrow \infty$,

$$Z_\nu = \underline{O}(\log \nu) \quad \text{a.s.}$$

Plugging this into (3.10) and using a familiar argument shows that for any

$\delta > 0$

$$|D_3(n) - D_4(n)| = O(\lambda_n^{-\delta}) \quad \text{a.s.}$$

as $n \rightarrow \infty$. This completes the proof of (3.6).

Now we show that if $s_n = \pi^{-1} \lambda_n^{-1/2m}$, then for some $C \in (0, \infty)$

$$\overline{\lim}_n (s_n \log \log s_n)^{-1/2} V(s_n) = C \quad \text{a.s.} \quad (3.11)$$

Since $\log \log s_n = \log \log (\pi^{-1} \sigma^{-1/m} n^{1/2m}) \sim \log \log n$, the lemma will follow from (3.6) and (3.11). (Note that the corresponding \lim in (3.11) will be $-C$ by symmetry.) Put

$$Y(u) = e^{-1/2u} V(e^u) \quad , \quad -\infty < u < \infty$$

$$u_n = \log s_n = A + (2m)^{-1} \log n \quad , \quad n = 1, 2, \dots$$

where $A = -\log \pi \sigma^{1/m}$ is a constant. One can check that $Y(u)$ is a stationary mean zero Gaussian process with covariance

$$\Gamma(u) = e^{-1/2u} \int_0^\infty G(e^{-u}t)G(t)dt$$

$$G(t) = h(t^m).$$

Equation (3.11) will follow by showing

$$\overline{\lim}_n (\log u_n)^{-1/2} Y(u_n) = 2^{1/2} \Gamma(0) \quad \text{a.s.} \quad (3.12)$$

To prove this latter we will apply Case 2 of Theorem D of Qualls (1977)

with $\alpha = 2$. Assuming this result applies, it states that for $c > 0$,

$$P[Y(u_n) > \Gamma(0)c(\log u_n)^{1/2}] = 0 \text{ or } 1$$

according as the integral

$$\int_1^\infty t^{-c^2/2} (\log t)^{c^2/4} (\log t - \frac{1}{2} \log \log t)^{-1/2} dt < \infty \text{ or } = \infty.$$

This latter integral converges for $c > 2^{1/2}$ and diverges for $c < 2^{1/2}$, which establishes (3.12).

In order to apply the aforementioned result in Qualls (1977), it is necessary to check that Y has continuous paths, $\Gamma(u) = o(1/\log u)$ as $u \rightarrow \infty$, and $\Gamma''(0) < 0$. To show Y is a.s. continuous, it suffices to show that V is a.s. continuous on $[0, \infty)$, and for this it suffices to show V is a.s. continuous on any finite interval $[0, b]$ and then let $b \rightarrow \infty$ through some countable set. Now let A be an event of probability 1 on which W is continuous and for which there exists T_0 (depending on the path of W) such that for all $T \geq T_0$

$$(2T \log \log T)^{-1/2} \sup_{\substack{0 \leq t \leq T/2 \\ 0 \leq s \leq T/2}} |W(t+s) - W(t)| \leq 2. \quad (3.13)$$

Such an event A exists by equation (1.2.4), page 30 of Csörgö and Révész

(1981). Let $\varepsilon > 0$ be given. Pick $T_1 \geq T_0$ such that

$$4 \int_{T_1/2b}^{\infty} 4(bt \log \log bt)^{1/2} |g(t)| dt < \varepsilon/2. \quad (3.14)$$

This is possible by (3.5). Now

$$\begin{aligned} \sup_{0 \leq s_1 \leq s_2 \leq b} |V(s_1) - V(s_2)| &\leq \int_{T_1/2b}^{\infty} \sup_{0 \leq s_1 \leq s_2 \leq b} |W(s_1 t) - W(s_2 t)| |g(t)| dt \\ &+ \int_0^{T_1/2b} \sup_{0 \leq s_1 \leq s_2 \leq b} |W(s_1 t) - W(s_2 t)| |g(t)| dt = I_1 + I_2. \end{aligned}$$

Since

$$\sup_{0 \leq s_1 \leq s_2 \leq b} |W(s_1 t) - W(s_2 t)| \leq \sup_{\substack{0 \leq u \leq bt \\ 0 \leq v \leq bt}} |W(u+v) - W(u)|$$

it follows from (3.13) and (3.14) that $I_1 < \varepsilon/2$. A straightforward uniform continuity argument shows that there is a $\delta > 0$ such that $|s_1 - s_2| \leq \delta$ implies $I_2 < \varepsilon/2$. This completes the proof of continuity.

To show $\Gamma(u) = o(1/\log u)$ as $u \rightarrow \infty$, we have for $u > 0$ and $\tau \neq \frac{1}{2}k - m$

$$\begin{aligned} \Gamma(u) &\leq e^{-(\frac{1}{2} + \tau/2m)u} \left[\int_0^1 t^{t/m} [1 + \tau^{1/2m}]^{-k} dt \right. \\ &\quad \left. + \int_1^{e^u} t^{(2\tau-k)/2m} dt + e^{ku/2m} \int_{e^u}^{\infty} t^{(\tau-k)/m} dt \right] \\ &= o(e^{-(\frac{1}{2} + \tau/2m)u}) + o(e^{((\tau-k)/2m + \frac{1}{2})u}). \end{aligned}$$

As all the exponents are negative, the desired result follows. If

$\tau = \frac{1}{2}k - m$, then the middle integral is treated differently but the result still holds.

To show that $\Gamma''(0) < 0$, first note that

$$\Gamma''(0) = \frac{1}{4} \int_0^{\infty} G^2(t) dt + \int_0^{\infty} [2tg(t) + t^2g'(t)]G(t) dt. \quad (3.15)$$

The calculation is carried out by differentiating under the integral sign which is justified by the dominated convergence theorem and by the fact that there exists a constant c such that for all $u \in [-1, 1]$, $|G(e^{-u}t)|$, $|tg(e^{-u}t)|$, and $|t^2g'(e^{-u}t)|$ are all bounded by $c \min\{1, t\}^{2m(\tau-k)}$, which is square integrable on $(0, \infty)$ by (3.5). Noting that the quantity in brackets on the r.h.s. of (3.15) is $d/dt[t^2g(t)]$, an integration by parts will yield

$$\Gamma''(0) = \frac{1}{4} \int_0^{\infty} G^2(t) dt - \int_0^{\infty} t^2g^2(t) dt.$$

Let I_1 denote the first integral in this last expression and I_2 the second.

Then integration by parts gives

$$I_1 = -2 \int_0^{\infty} tg(t)G(t) dt$$

and Cauchy-Schwarz applied to this integral gives $I_1 < 2I_1^{1/2}I_2^{1/2}$ which implies $I_1 < 4I_2$ and hence $\Gamma''(0) = \frac{1}{4}I_1 - I_2 < 0$. \square

Corollary 3.3. *Let $-1/2m < \rho < 1 - 1/2m$. Suppose that $\Delta_n > 0$ is chosen so that*

$$P[\|\hat{\beta}_n - \beta\|_\rho^2 \leq \Delta_n | Y_n] = 1 - \alpha \quad (3.16)$$

for some $\alpha \in (0,1)$. Then

$$\liminf_{n \rightarrow \infty} P[\|\hat{\beta}_n - \beta\|_\rho^2 \leq \Delta_n | \beta] = 0, \quad \text{a.s.} \quad (3.17)$$

Proof. By Theorem 3.1(a)

$$\Delta_n \sim \mu_n + z_\alpha \tau_n \quad (3.18)$$

where z_α is the upper $100\alpha\%$ point of the $N(0,1)$ distribution. By Theorem 3.1(b)

$$|P[\|\hat{\beta}_n - \beta\|_\rho^2 \leq \Delta_n | \beta] - \Phi((z_\alpha \tau_n - \Omega_n)/\eta_n)| \rightarrow 0$$

a.s. as $n \rightarrow \infty$, where Φ denotes the $N(0,1)$ distribution function. (Note that convergence in distribution to $N(0,1)$ implies uniform convergence of the distribution function to Φ .) Now τ_n/η_n tends to a finite limit and

$$\eta_n \approx \lambda_n^{1-\rho-1/4m}$$

by (3.3). Hence, by Lemma 3.2

$$\liminf_n (z_\alpha \tau_n - \Omega_n)/\eta_n = -\infty$$

which proves (3.16) \square

Corollary 3.4. *With the setup of the previous corollary, the prior probability that the $(1-\alpha)$ Bayesian posterior probability region $\{\theta \in \Theta_\rho: \|\hat{\beta}_n - \theta\|_\rho^2 \leq \Delta_n\}$ is conservative in its frequentist coverage*

probability tends to $\Phi(z_\alpha(1-\tau^{1/2})(1-\tau^2)^{-1/2})$ where

$$\tau^2 = (1-\rho-1/4m)(2\rho/3-1/6m).$$

Proof. We wish to calculate

$$P\{\beta: P[\|\hat{\beta}_n - \beta\|_p^2 \leq \Delta_n | \beta] \geq 1-\alpha\}.$$

The difference between this and

$$P[\Phi((z_\alpha \tau_n - \Omega_n)/\eta_n) \geq 1-\alpha] = P[\Omega_n \leq (\tau_n - \eta_n)z_\alpha]$$

tends to 0. One can show as in Theorem 3.1(a) that $\underline{L}(\Omega_n/(\tau_n^2 - \eta_n^2)^{1/2}) \Rightarrow N(0,1)$. Using this and (3.3) one can show that the last displayed probability tends to a limit of the form $\Phi(z_\alpha R)$ where R is an algebraic form in some beta functions. If one substitutes $B(x,y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$ and uses the recurrence relation $\Gamma(x+1) = x\Gamma(x)$, then the claimed result follows after some algebraic manipulation. \square

4. DISCRETE TIME ESTIMATION

In this section we consider the original discrete time problem in (1.1), where β is given the prior specified in (1.2). By virtue of the next result, it follows that Theorem 3.1 and Corollaries 3.3 and 3.4 hold in this setting as well, thus justifying the statements made in the introduction.

Theorem 4.1. *Suppose that $m \geq 2$ and*

$$-1/2m < \rho < 1 - 1/2m$$

and that

$$E |\varepsilon_i|^{2+\delta} < \infty$$

where

$$\delta \geq 6/(2m-3).$$

Then there exists a probability space carrying versions of β , $\varepsilon_1, \varepsilon_2, \dots$ and a sequence $\{\tilde{\varepsilon}_n(t): 0 \leq t \leq 1\}$ of Gaussian white noises such that if

$$Y_{ni} = \beta(i/n) + \varepsilon_i, \quad 1 \leq i \leq n$$

$$\hat{\beta}_n = E[\beta | Y_n], \quad \delta_n = \hat{\beta}_n - \beta$$

$$\tilde{Y}_n(t) = \beta(t) + \sigma n^{-1/2} \tilde{\varepsilon}_n(t), \quad 0 \leq t \leq 1$$

$$\tilde{\beta}_n = E[\beta | \tilde{Y}_n], \quad \tilde{\delta}_n = \tilde{\beta}_n - \beta$$

then

$$\|\delta_n - \tilde{\delta}_n\|_\rho^2 = o_p(\lambda_n^{1-\rho}).$$

Remarks. Note that $\|\tilde{\delta}_n\|_\rho^2 = \underline{O}_p(\lambda_n^{1-\rho-1/2m})$ follows from $E\|\tilde{\delta}_n\|_\rho^2 = \mu_n(\rho)$.

Since $\|\delta_n - \tilde{\delta}_n\|_\rho^2 = \underline{O}_p(\lambda_n^{1-\rho-1/2m})$, it follows that $\|\delta_n\|_\rho^2 = \underline{O}_p(\lambda_n^{1-\rho-1/2m})$.

Hence

$$\begin{aligned} \left| \|\delta_n\|_\rho^2 - \|\tilde{\delta}_n\|_\rho^2 \right| &\leq \|\delta_n - \tilde{\delta}_n\|_\rho [\|\delta_n\|_\rho + \|\tilde{\delta}_n\|_\rho] \\ &= \underline{O}_p(\lambda_n^{(1-\rho)/2}) \underline{O}_p(\lambda_n^{(1-\rho-1/2m)/2}) = \underline{O}_p(\lambda_n^{1-\rho-1/4m}). \end{aligned}$$

Now $\tau_n \approx \lambda_n^{1-\rho-1/4m}$ and $\eta_n \approx \lambda_n^{1-\rho-1/4m}$, so we may replace the continuous time estimation error (now denoted $\tilde{\delta}_n$) with the discrete time estimation error δ_n in Theorem 3.1, and the results still hold. Corollaries 3.3 and 3.4 depend only on these results and the properties of Ω_n , which depends only on β .

Proof. We have

$$\|\delta_n - \tilde{\delta}_n\|_\rho^2 \leq 2\|\tau_n - s_n\|_\rho^2 + 2\|(B_n - \tilde{B}_n)\beta\|_\rho^2$$

where

$$\tau_n = \hat{\beta}_n - E[\hat{\beta}_n | \beta], \quad s_n = \tilde{\beta}_n - E[\tilde{\beta}_n | \beta]$$

$$B_n\beta = E[\hat{\beta}_n - \beta | \beta] = [(\lambda_n I + U_n)^{-1} U_n - I]\beta$$

$$\tilde{B}_n\beta = E[\tilde{\beta}_n - \beta | \beta] = [(\lambda_n I + U)^{-1} U - I]\beta.$$

We will show that each of the two quantities on the r.h.s. of the inequality above has the correct order as $n \rightarrow \infty$.

Now

$$\tau_n = (\lambda_n I + U_n)^{-1} X_n^* \underline{\varepsilon}_n$$

and following the notation of Cox (1984b) put

$$\bar{\tau}_n = (\lambda_n I + U)^{-1} X_n^* \varepsilon_n.$$

Now we apply some results in section 4 of Cox (1986). One can show that for all x_1, x_2 in Θ

$$|\langle (U - U_n)x_1, x_2 \rangle_\Theta| \leq Cn^{-1}[\|x_1\|_{1/m}\|x_2\|_0 + \|x_1\|_0\|x_2\|_{1/m}]$$

where $C \in (0, \infty)$ is a constant, so Assumption 4.1(f) of Cox (1986) holds with $k_n = Cn^{-1}$, $s = 1/m$, $j = 2$. See the remarks after this Assumption for a proof of the above inequality. Now by equation (4.6) of Cox (1986),

$$E\|\tau_n - \bar{\tau}_n\|_\rho^2 \leq Cn^{-2}[C_2(\lambda_n, \rho + 1/m)E\|\tau_n\|_0^2 + C_2(\lambda_n, \rho)E\|\tau_n\|_{1/m}^2].$$

This latter inequality can be inverted for $\rho = 0, 1/m$ to give bounds on $E\|\tau_n\|_0^2$ and $E\|\tau_n\|_{1/m}^2$ in terms of $E\|\bar{\tau}_n\|_0^2$ and $E\|\bar{\tau}_n\|_{1/m}^2$. When these are substituted back into this inequality and the estimates on C_2 are used there results

$$E\|\tau_n - \bar{\tau}_n\|_\rho^2 \leq Cn^{-3}\lambda_n^{-\rho-2/m}.$$

It is easily seen that this last quantity is $\underline{o}(\lambda_n^{1-\rho})$ as $m \geq 2$.

Now put

$$\bar{s}_n = \sigma n^{-1/2} \int_0^1 (\lambda_n I + U)^{-1} \xi(t) dB_n \circ F_n(t)$$

where $F_n(t) = n^{-1}[nt]$ as in (3.7) of Cox (1984), then following (3.9) of that paper,

$$\|\bar{\tau}_n - \bar{s}_n\|_\rho^2 = n^{-1}C_2(\lambda_n, \rho + 1/m)\underline{o}(n^{-\delta/(2+\delta)})$$

a.s. Using the estimate on C_2 , the definition of λ_n , and the hypothesis that $\delta \geq 6/(2m-3)$, it follows that the last displayed quantity is $\underline{o}(\lambda_n^{1-\rho})$, a.s.

Next, we have from (3.10) of Cox (1984b) that

$$\|\bar{s}_n - s_n\|_\rho^2 = \underline{O}_p(n^{-1}(n^{-1} \log n) C_2(\lambda_n, \rho + 1/m))$$

and this is $\underline{o}_p(\lambda_n^{1-\rho})$. This completes the proof that $\|\tau_n - s_n\|_\rho^2 = \underline{o}_p(\lambda_n^{1-\rho})$.

Finally we must take care of $\|(B_n - \tilde{B}_n)\beta\|_\rho^2$ where B_n and \tilde{B}_n denote the bias operators for the respective problems. Let $\max\{\rho, 1/m\} < \tau < 1 - 1/2m$, and let $\|\cdot\|_{\tau, \rho}$ denote the operator norm of a linear operator from Θ_τ to Θ_ρ . Then by equation (4.6) of Cox (1986),

$$\begin{aligned} \|B_n - \tilde{B}_n\|_{\tau, \rho} &\leq C n^{-1} \{C_2^{1/2}(\lambda_n, \rho + 1/m) \|\tilde{B}_n\|_{\tau, 0} + \\ &\quad C_2^{1/2}(\lambda_n, \rho + 1/m) \lambda_n^{1/2m} \|\tilde{B}_n\|_{\tau, 1/m} + \\ &\quad C_2^{1/2}(\lambda_n, \rho) \lambda_n^{-1/2m} \|\tilde{B}_n\|_{\tau, 0} + C_2^{1/2}(\lambda_n, \rho) \|\tilde{B}_n\|_{\tau, 1/m}\}. \end{aligned}$$

Now by Theorem 2.3(c) of Cox (1986), $\|\tilde{B}_n\|_{\tau, \rho} \leq C \lambda^{(\tau-\rho)/2}$ as long as $\rho \leq \tau \leq \rho+2$. Plugging this into the above expression and using the estimates on C_2 yields

$$\|B_n - \tilde{B}_n\|_{\tau, \rho} \leq C \{\lambda_n^{2-\rho-3/2m+\tau}\}^{1/2}.$$

Now

$$\|(B_n - \tilde{B}_n)\beta\|_\rho^2 \leq C \lambda_n^{2-\rho-3/2m+\tau} \|\beta\|_\tau^2.$$

As $\tau < 1 - 1/2m$, $\|\beta\|_\tau^2 < \infty$ a.s., by (3.2) and Proposition 2.5(f), and so the last expression is a.s.

$$\underline{O}_p(\lambda_n^{2-\rho-3/2m+r}) = \underline{o}_p(\lambda_n^{1-\rho}). \quad \square$$

5. CONCLUDING REMARKS

To provide more insight into the foregoing analysis, consider the case where $\dim \Theta = k < \infty$. Then it is easy to obtain the analog of Theorem 3.1. Calculating the asymptotic posterior distribution of $\|\delta_n\|_\rho^2$, we have

$$\underline{L}(\|\delta_n\|_\rho^2) = \underline{L} \left[\sum_{\nu=1}^k \lambda_n \gamma_\nu^\rho (1 + \lambda_n \gamma_\nu)^{-1} \varepsilon_\nu^2 \right].$$

Since $(1 + \lambda_n \gamma_\nu)^{-1} \rightarrow 1$ as $n \rightarrow \infty$ for $\nu = 1, 2, \dots, k$, it follows that

$$\underline{L}(\lambda_n^{-1} \|\delta_n\|_\rho^2) \Rightarrow \underline{L} \left[\sum_{\nu=1}^k \gamma_\nu^\rho \varepsilon_\nu^2 \right]$$

where the distribution on the right is a weighted sum of χ^2 random variables. The corresponding calculations for the sampling distribution are

$$\begin{aligned} \underline{L}(\|\delta_n\|_\rho^2 | \beta) = \underline{L} \left[\lambda_n \sum_{\nu=1}^k \gamma_\nu^\rho (1 + \lambda_n \gamma_\nu)^{-2} \varepsilon_\nu^2 \right. \\ \left. + \lambda_n^{3/2} \sum_{\nu=1}^k \gamma_\nu^{1/2+\rho} (1 + \lambda_n \gamma_\nu)^{-2} \varepsilon_\nu \beta_\nu + \lambda_n^2 \sum_{\nu=1}^k \gamma_\nu^{1+\rho} (1 + \lambda_n \gamma_\nu)^{-2} \beta_\nu^2 \right]. \end{aligned}$$

The second and third summations on the r.h.s. (which result from the bias) are each $o_p(\lambda_n^{-1})$, so

$$\underline{L}(\lambda_n^{-1} \|\delta_n\|_\rho^2 | \beta) \Rightarrow \underline{L} \left[\sum_{\nu=1}^k \gamma_\nu^\rho \varepsilon_\nu^2 \right].$$

Thus, the asymptotic posterior and sampling distributions are identical, so a $(1-\alpha)$ posterior probability region of the form in (3.16) has (asymptotically) the right coverage probability from the sampling point of view. This results from the fact that the random (conditional on β) variability

dominates both the sampling and posterior distributions, with the bias $E[\hat{\beta}_n - \beta | \beta]$ being of smaller order.

In the case when $\dim \Theta = \infty$, the bias and random variability are always of the same order of magnitude. This is the driving force behind Corollaries 3.3 and 3.4.

Concerning extensions of the major results (Theorem 3.1, Lemma 3.2, and Theorem 4.1), one would like to look at more general Gaussian priors and more general quantities than $\|\delta_n\|_\rho^2$. I believe that analogous results can be obtained (for $\|\delta_n\|_\rho^2$ in any setting wherein $\gamma_\nu \approx \nu^\tau$ for some $\tau > 0$ (see Cox (1986)). The specific setting here allowed us to use the sharp eigenvalue asymptotics (3.3) at some points in Lemma 3.1 (see (3.8)). The extension to other quadratic forms can also probably be accomplished. As in Wahba (1983), we would like to construct a confidence interval for $\beta(t)$ for some $t \in [0,1]$. In the setup of section 3, note that

$$\underline{L}(\delta_n(t)) = N(0, \lambda_n \sum_\nu (1 + \lambda_n \gamma_\nu)^{-1} \varphi_\nu^2(t))$$

$$\underline{L}(\delta_n(t) | \beta) = N(M_n, \lambda_n \sum_\nu (1 + \lambda_n \gamma_\nu)^{-2} \varphi_\nu^2(t))$$

where

$$M_n = \lambda_n \sum_\nu \gamma_\nu^{1/2} (1 + \lambda_n \gamma_\nu)^{-1} \beta_\nu \varphi_\nu(t).$$

I conjecture that both variances above are $\approx \lambda_n^{1-1/2m}$, and also that

$$\limsup_{n \rightarrow \infty} [\lambda_n^{(1-1/2m)} (\log \log 1/\lambda_n)]^{-1/2} M_n \in (0, \infty) \quad , \quad \text{a.s.}$$

This would imply analogs of Corollaries 3.3 and 3.4 for the "confidence"

intervals for $\delta_n(t)$. The difficulty here is in obtaining results on the behavior of $\varphi_\nu(t)$.

References

- Cox, D. (1984a). Multivariate smoothing spline functions. *SIAM J. on Numer. Anal.* **21**, 789-813.
- Cox, D. (1984b). Gaussian approximation of smoothing splines. Submitted to *Z. Wahrscheinlichkeitstheorie*.
- Cox, D. (1986). Approximation of method of regularization estimators. Submitted to *Ann. Statist.*
- Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Math.* **31**, 377-403.
- Csörgö, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, New York.
- Gradshteyn, I. and Ryzhik, I. (1965). *Tables of Integrals, Series, and Products*. Academic Press, New York.
- Kallianpur, G. (1971). Abstract Wiener processes and their reproducing kernel Hilbert spaces. *Z. Wahrscheinlichkeitstheorie* **17**, 113-123.
- Kallianpur, G. and Karandikar, R. (1985). White noise calculus and non-linear filtering theory. *Ann. Probability* **13**, 1033-1107.
- Kimeldorf, G. and Wahba, G. (1970a). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Stat.* **41**, 495-502.

- Kimeldorf, G. and Wahba, G. (1970b). Spline functions and stochastic processes. *Sankhya A* **32**, 173-180.
- Kuelbs, J. (1970). Gaussian measures on a Banach space. *J. Func. Anal.* **5**, 354-367.
- Kuelbs, J. (1971). Expansions of vectors in a Banach space related to Gaussian measures. *Proc. Amer. Math. Soc.* **27**, 364-370.
- Kuo, H.H. (1975). *Gaussian Measures on Banach Spaces*. Lectures Notes in Mathematics No. 463, Springer-Verlag, Berlin.
- Naimark, M.A. (1967). *Linear Differential Operators*. Frederick Ungar, New York.
- Qualls, C. (1977). The law of the iterated logarithm on arbitrary sequences for stationary Gaussian processes and Brownian motion. *Ann. Probability* **5**, 724-739.
- Speckman, P. (1984). The asymptotic integrated mean square error for smoothing noisy data by splines. To appear in *Numerische Math.*
- Speckman, P. (1983). Efficient nonparametric regression with cross validated smoothing splines. Submitted to *Ann. Statist.*
- Wahba, G. (1978). Improper priors, spline smoothing, and the problem of guarding against model errors in regression. *J. Royal Statist. Soc. B* **40**, 364-372.
- Wahba, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline. *J. Royal Statist. Soc. B* **45**, 133-150.

Wecker, W. and Ansley, C. (1983). The signal extraction approach to non-linear regression and spline smoothing. *J. Amer. Statist. Assoc.* 78, 81-89.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 93	2. GOVT ACCESSION NO. AD-A176 788	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) AN ANALYSIS OF BAYESIAN INFERENCE FOR NON-PARAMETRIC REGRESSION		5. TYPE OF REPORT & PERIOD COVERED TR 12/1/85-11/30/86
7. AUTHOR(s) Dennis D. Cox		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics, GN-22 University of Washington Seattle, WA 98195		8. CONTRACT OR GRANT NUMBER(s) N00014-84-C-0169
11. CONTROLLING OFFICE NAME AND ADDRESS ONR Code N63374 1107 NE 45th St Seattle, WA 98105		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-661-003
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE October 1986
		13. NUMBER OF PAGES 38
		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Bayesian inference, nonparametric regression, confidence regions, signal extraction, smoothing splines		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)		

END

3-87

Dtic